

ニューラルネットワークの自己生成入力データによる 時系列データの予測 (河川流量の場合)

三重野 保男[†] 有馬 明広^{††}

Prediction of Time Series Data by Neural Networks with Self-Generating Input Data
(In Case of the Flow of the River)

MIENO Yasuo ARIMA Akihiro

Estimation of time series data has been studied by various methods including neural networks in recently. It is difficult to estimate the time series data in long period of time span. Therefore, this paper describes a new method to predict the time series data of long period by using "self-generating input data". After tuning the synaptic weights that was derived by the learning of neural networks, our method returns to the calculated output data of neural networks to the input side as input data for next calculating time. We verified the validity of our method by forecasting for the periodic function including the attenuation clause. Finally, we estimated the flow of the river in long period of time span in good precision. Consequently, our proposed method is ascertained validity.

Keywords : Prediction of Time Series Data, Neural Network, Self-Generating Input Data,
Flow of the River

1. まえがき

時系列データの予測は、現在までにニューラルネットワークを用いた方法も含め、数多くの方法で実行されてきている。しかしながらいずれもその予測期間は短く、長期間の予測は困難なためあまり見かけない。そこで本論文では学習終了後に予測によって得られた値を入力へ戻し、それによって新たに次の予測値を得るという“自己生成入力データ”による予測方法をとることによって長時間の時系列データの予測を行うことにした。本論文での予測手法を評価するために減衰項を含んだ周期関数で予測を行い、その有効性を検証し、次に肝属川の流量予測を行った。

一般に雨は地面に降ったあと大部分は地下に浸透し、またはそのまま地表を流れるなどして地下水や河川の流量となる。その時どの程度の量が地下に浸透してい

くのか、そして浸透した雨はどのくらいの間地面に貯えられて出てくるのかなど、現状ではその過程を把握することは、雨量と流量の間に強い非線形性がみられるため極めて困難である。これを解決する方法の一つとして非線形近似を得意とするニューラルネットワークによる流量解析および予測があげられる。本論文では、ニューラルネットワークによって過去の雨量や過去の流量から、雨が降ってからそれが流量に反映される過程を学習し、自己生成入力データによる長期間の流量予測を行った。

2. 学習および予測の手法

時系列データを予測する手法としては、ニューラルネットワーク処理の基本機能である時系列データの特徴抽出機能を用いる。与えられた時系列データを X_n ($n=1, 2, \dots$) とする。図1に示すように入力層のニューロンの個数を α 、出力層のニューロンの個数を β 、そして学習パターンを i ($i=1, 2, \dots$)

[†]鹿児島工業高等専門学校情報工学科

^{††}電気情報システム工学専攻科

とし、ニューラルネットワークの入力を $X_i \sim X_{\alpha+i}$ 、教師データとして $X_{\alpha+i+1} \sim X_{\alpha+i+\beta}$ を与えるようなモデルを考える。つまり α 個の入力データと $\alpha+1$ 以降の β 個の教師データとの関係をパターンごとに学習していくことになる。最終的にはこのようにして学習したニューラルネットワークを用いて予測を行っていく。なお、学習法はシグモイド関数を用いた誤差逆伝播法⁹⁾を用いている。なお本論文での予測誤差は、関数の場合と流量の場合を比較出来るよう、正規化した範囲内でどの程度のものかという意味で、関数の予測の場合は誤差の絶対値の平均に100をかけて単位をパーセントとし、流量の場合は対数をとっているのでもに反した後全データを最大値で割り、関数と同じようにパーセントで計算している。

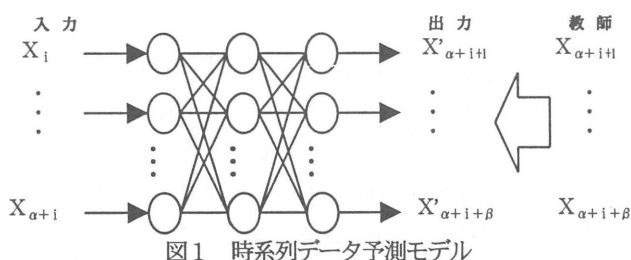


図1 時系列データ予測モデル

3. 自己生成入力データによる関数の予測の検証

出力層にニューロンが1つしかないニューラルネットワークでは、本来なら1時刻先の流量しか予測することはできない。しかし得られた1時刻分の予測流量データを入力側に戻すことによってさらに1時刻分の予測を行うことができる。この作業を繰り返すことにより1時刻だけの予測ではなく複数時刻先の流量予測を行うことにした。

まず、その有効性を確かめるために式(1)に示した簡単な減衰項を含んだ関数の予測を行い検証した。

$$y = \frac{1}{e^{x/100}} * \sin\left(\frac{\pi x}{18}\right) \quad (1)$$

x は(0, 1, 2, ..., n)

図2はこの関数を対象にして学習し、予測を行った結果である。このときのネットワークの構成は出力層1個、中間層6個、入力層12個でフィードバックのあるモデルになっている。学習係数は0.5で慣性項係数は0.4、学習回数は13000回である。

ニューラルネットワークによる予測値は少なからず誤差を含んでいるもので、この方法ではその誤差を含んだ値を入力に用いることになるが、学習がうまく出来ている場合にはその誤差の影響をあまり受けることなく次の予測値を得ることができる。

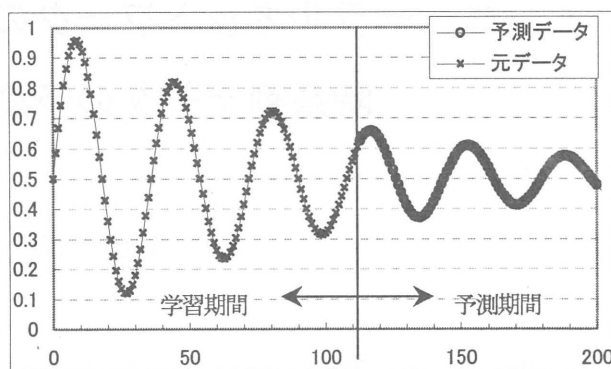


図2 関数での予測結果

学習はデータの0から1時刻ずつずらしていった100組のパターンを学習し、その先をこの方法を用いて87時刻予測したものだが、予測部分の誤差は0.2948[%]となっており、その予測値は元データとよく対応している。この結果より、我々の提案する方法を用いることで時系列データの複数時刻の予測が可能であるとして肝属川の流量予測を行い、その結果を検証することにした。

4. 肝属川流域での流量予測システムの構築

4.1 入力データ

(1) 使用データ

河川の流量を予測するにあたって必要なデータは、まず過去に降った雨量と教師データとして使用する流量のデータの二つが考えられる。河川の流量は主に過去の雨量に大きく影響を受けるため入力側には雨量を用いることが考えられるからである。しかし河川の流量を予測するとき、入力側に過去の雨量だけでなく過去の流量のデータを加えることで予測結果がよくなるのが予想される。そこで、研究に使用するデータとして流量と雨量のデータを用意した。

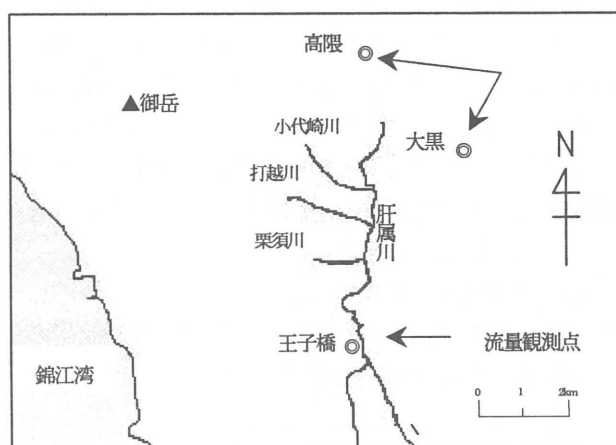


図3 肝属川流域図

予測の対象としたのは図3に示した鹿児島県鹿屋市にある肝属川の流量で、データは王子橋における流量のデータと、肝属川の上流域にあたる大黒と高隈の雨量観測所での雨量データを使用した。流量予測の対象として肝属川を選んだのは、ダムがなく自然の形の流量データが得られているためである。学習データとして10年以上も前からのあまりに遠い過去データを用いることは、河川改修等で状況が変化することを考えれば現状を表現するよい学習結果が得られないと考えた¹⁾。そこで本論文では日単位でのデータを1993年から1995年まで用意し、さらに一時間あたりの流量、雨量のデータを1995年の4月1日の午前0時から5月29日の午前9時までの1400時間分を用意した。これら正規化前の使用データのグラフを一部拡大して図4-1と図4-2に示す。

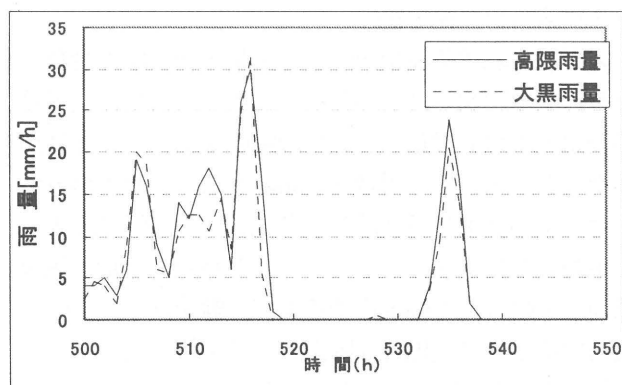


図4-1 雨量データ(一部)

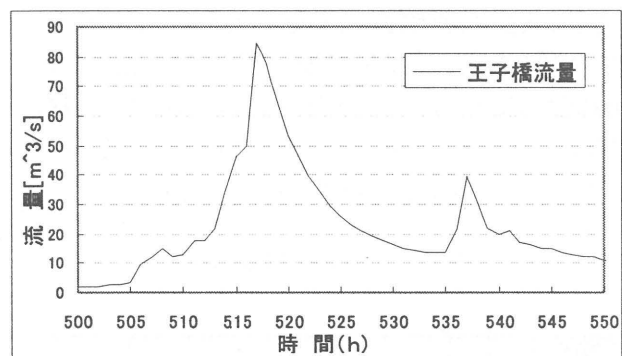


図4-2 流量データ(一部)

(2) データの基準化

ニューラルネットワークでの使用のため、入力雨量データは最大値で基準化した値を用い、入力および教師データとして使用する流量データは最大流量値を底にとった対数で全データを基準化した。これは流量データの流量が大きい時期のデータと平常時の流量データの差が大きく、最大値でそのまま基準化してしまうと学習データの分布に偏りが生じてしまうためである³⁾。なお、ニューラルネットワークへのデータの与え

方にはデータをそのまま（絶対値）与える方法とデータの変化分（差分）を与える方法があり、流量予測や水位予測にニューラルネットワークを用いる場合、絶対量を用いる方法では流量・水位の大きいところに誤差が表れやすく、差分にすることにより精度が向上するとの報告がある²⁾。

今回は入力に基準化した流量をそのまま与える場合と差分で与える場合を比較するために差分の入力データも用意した。差分データは、基準化した流量データをもとに差分をとり、さらにその最大値が1、最小値が0となるようにで基準化したものを使用した。正規化後の使用データを図5に示す。正規化雨量は、見にくいが大黒のものと高隈の雨量が重なっている。

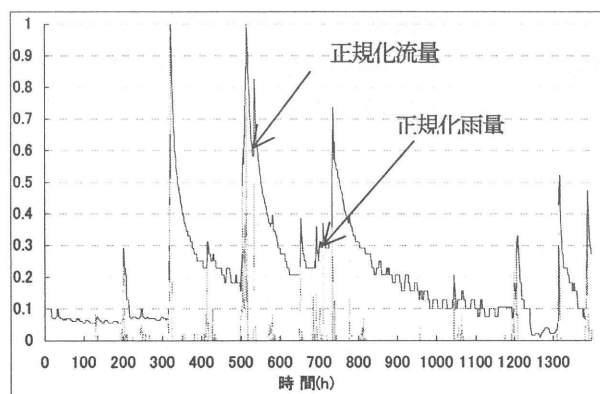


図5 正規化後の使用データ

4-2 流量予測モデルの構築

(1) 学習・予測条件

河川流量予測に用いるデータは前項で紹介した流量および雨量データである。このうち最初の500時間を学習用とし、そのあとの900時間分を予測用として用いた。予測に用いるニューラルネットワークの結合の重みは学習終了時のまま固定し、1時間毎に予測を行った。表1にニューラルネットワークの基本パラメータを示す。ここで、学習係数、各素子数などは予備実験より決定したものである。これから先はこのパラメータを使用して計算をしている。

表1 基本パラメータ

項目	値
学習係数	0.5
慣性項係数	0.85
学習回数	10000
中間層仮素子数	6

(2) 出力層の個数

本研究の目的は、流量の予測であるが、出力層素子の個数は一回の予測でどれだけ先まで予測できるの

かということに大きく関わってくる。例えば出力層素子 n 個の場合、予測時に 1 パターンの入力データをニューラルネットワークに与えたときに、 n 個の予測結果が得られるということである。今、与えられた時系列データが x_i ($i = 1, 2, \dots$) とすると、一般的に複数時刻の予測を行う場合、入力を $x_1 \sim x_n$ 、教師として $x_{n+1} \sim x_N$ とするのが一般的である。長期予測の方法としては出力層の素子数を増やすことが考えられるが、この場合ではデータ間の写像を学習するので時間要素の入る余地は無い。このような場合でも、ある時間区間のデータを一つのパターンとみなし静的に考えるならば既存の方法でも処理できるが、重要な時間相関が無視される。このような構成のネットワークでは学習用入力および教師の両データの写像を記憶するので、 x_{n+1} は別にしても、それ以降の $x_{n+2} \sim x_N$ は関連する 1 あるいは数時刻前のデータの影響が無視され、時系列の特質は失われる。⁷⁾

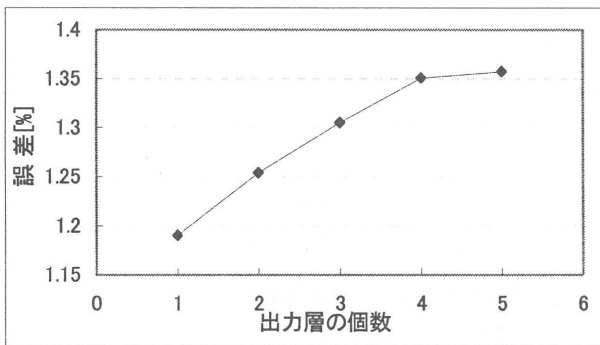


図6 出力層素子数と誤差

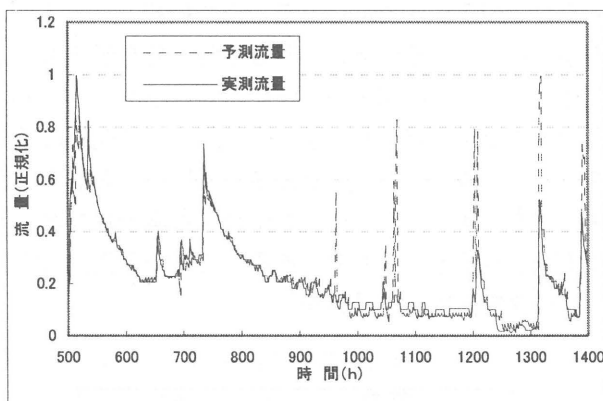


図7 出力層5つのときの予測結果

図6は出力層の個数のみを変化させて誤差を比較したもののだが出力層の個数が増加するにつれ誤差が大きくなっているのが分かる。図7は出力層が複数の場合の一例として出力層が5個の場合の計算結果を示したものである。この図をみると予測部分の特に後半の部分でかなり大きな誤差が表れているのが分かる。

この結果より、肝属川の流量予測モデルは図8のように出力層の素子数を1個に固定し、入力層、中間層の個数を変化させ、流量予測に最適な構成のニューラルネットワークを構築していく。

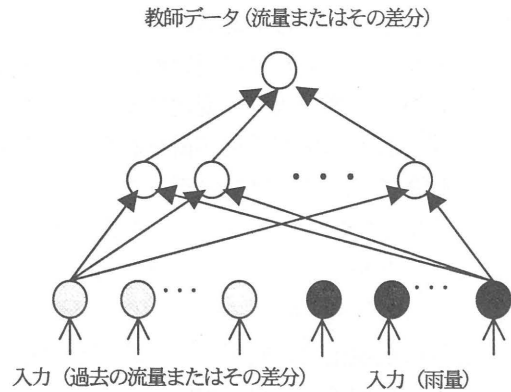


図8 流量予測モデルの概念図

(3) 入力データの決定

入力層における流量データと雨量データ組み合わせとその個数の割合について検討した。入力層に王子橋の流量と大黒の雨量、そして高隈の雨量という3種類のデータを入力する場合を Case-1 とし、その組み合わせによる結果を予測誤差[%]によって比較した。その結果を表2に示す。このとき大黒と高隈の雨量の個数は同じものとして個数をそろえて実験を行った。次に、入力に流量及び大黒と高隈の雨量を平均した値を入力とした場合を Case-2 として計算を行い、その結果を表3に示す。表2、3より入力層の組み合わせでもっともよい組み合わせは、Case-2 の入力素子数が流量2個、雨量3個という結果であった。この結果より複数の観測点での雨量データを使用する場合には入力層に並列に入力するよりも、データを単純平均したものを入力した方が予測精度はあがることが分かった。

表2 Case-1 の計算結果

		流 量 →					
雨 量 ↓	個数	1	2	3	4	5	6
	1	0.6956	0.3996	0.396	0.399	0.4	0.5111
	2	0.3888	0.3812	0.383	0.387	0.408	0.4542
	3	0.3764	0.4009	0.429	0.4	0.4	0.4811
	4	0.4699	0.3857	0.379	0.388	0.399	0.489
	5	0.3814	0.3738	0.378	0.448	0.388	0.4598
	6	0.629	0.5214	0.438	0.418	0.476	0.468

単位：絶対平均誤差[%]

この他に、流量データの代わりに流量データの差分を使用することにより、予測精度の向上をこころみた。検証の手順としては流量の絶対量を与えていた場合と

条件を同じにし、同じように入力層での最適なデータの組み合わせと個数の割合を試したが、どの組み合わせもあまりよい結果を得ることが出来なかった。

これらの結果より、今回の事例の場合には入力および教師データには差分を与えるのではなく、流量の絶対量を用いた方がよいことがわかった。

表3 Case-2の計算結果

流量 →		1	2	3	4	5	6
個数							
雨量 ↓	1	0.3803	0.4163	0.427	0.385	0.409	0.4838
	2	0.3555	0.3618	0.352	0.383	0.381	0.4652
	3	0.35	0.3417	0.37	0.422	0.443	0.6049
	4	0.3585	0.3654	0.376	0.415	0.413	0.481
	5	0.3588	0.3618	0.392	0.388	0.4	0.509
	6	0.3835	0.4417	0.59	0.441	0.513	0.4286

単位：絶対平均誤差%

(4) 中間層の素子数および予測結果

前節のように、入力層に入力するデータの種類とその最適な組み合わせが決定したところで、次は基本パラメータとして仮に設定していた中間層の素子数を、最適な個数にするために中間層素子数を2～10まで変化させて誤差を計算した。その結果は図9のグラフに示すように、一番誤差が少ないのは中間層素子数が3個の場合である。

以上で肝属川の河川流量を予測するための最適なニューラルネットワークの各層の素子数および入力データの種類と組み合わせが決定したことになる。この構成のネットワークで予測した結果、誤差は0.340710[%]であった。

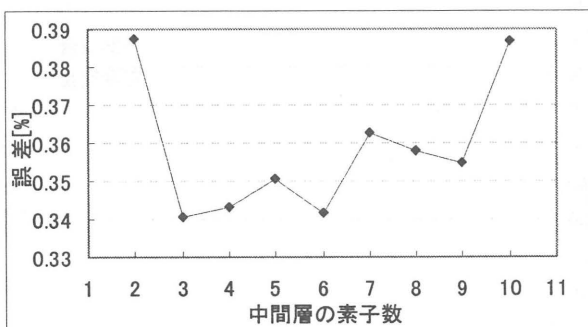


図9 中間層素子数と誤差

4-4 フィードバックを加えた場合（松葉モデル⁷⁾）

通常の多層ニューラルネットワークは入力と教師との写像関係を学習するので、時間的要素に入る余地はない。時系列データを対象とする場合、ある区間のデータを1パターンとみなし静的に考えるならば一般的

なネットワークでも処理できるが、重要な時間相関が無視され、予測などに用いる時にはあまりよい結果が期待できない。そこで時系列データの予測もモデルの一つである松葉モデル（フィードバックモデル）の要素をこれまでのネットワークに加えるとさらなる予測精度の向上につながる可能性があると考えて、図10のようなモデルを元にして予測を行った。

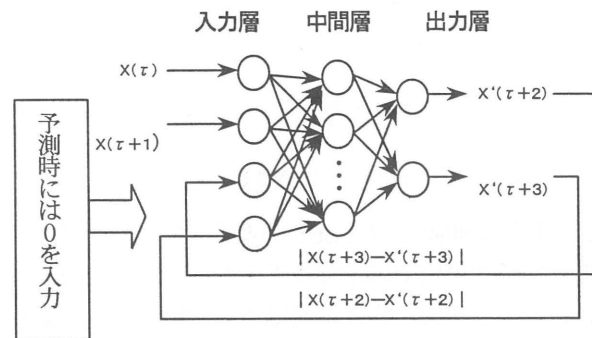


図10 松葉モデルの概念図

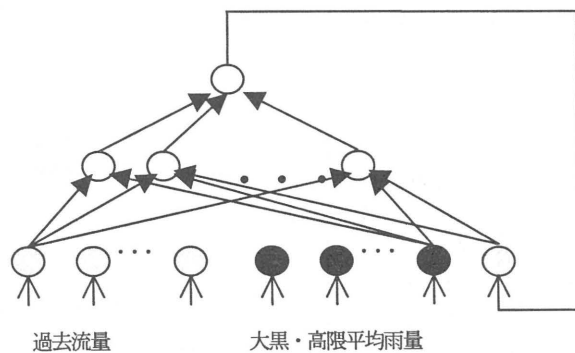


図11 流量予測のフィードバックモデル

このモデルは、学習中に教師データと出力との差の絶対値を、入力層に新たに設けた素子にフィードバックしていくというもので、学習終了後、予測時にはその新たな素子には0を入力して計算するというものである。このときの学習条件は前節までに決定したパラメータを使用している。そして結果としては、予測誤差は0.338756[%]となり(4)の結果である0.340710[%]よりも小さく、予測精度を上げることができた。予測結果のグラフを図12に示す。

(5) 日単位での流量予測

用意しておいた日単位のデータを使って前項までと同じ手順で流量予測の検証を行ったが、データが時間単位のものよりも荒くなり、学習も予測もどちらもうまくいかなかった。日単位での流量予測で一番精度の高かった予測結果のグラフを図13-1と図13-2に示す。このときは300(日)まで学習していて、そ

のあとの部分が予測部分である。

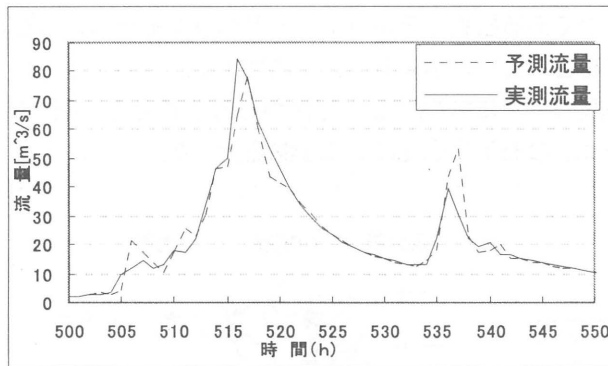


図12 時間単位での予測結果(一部)

この結果予測誤差は1.368550[%]となっていて、これまでの時間単位での予測結果よりもかなり誤差が大きいことが分かる。やはり流量予測の場合には、より細かく正確なデータであるほうが精度のよい予測が行えるということが分かった。このことより今後は時間単位での予測を対象として研究を行った。

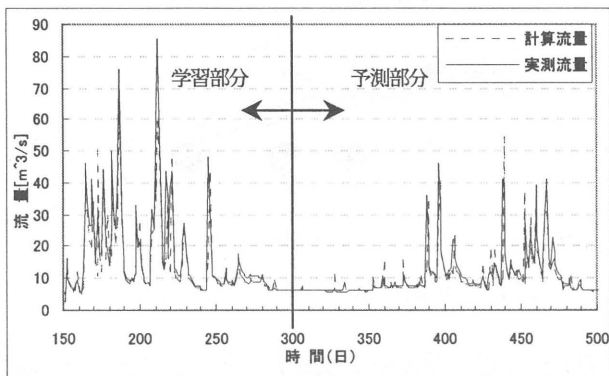


図13-1 日単位での計算結果(一部)

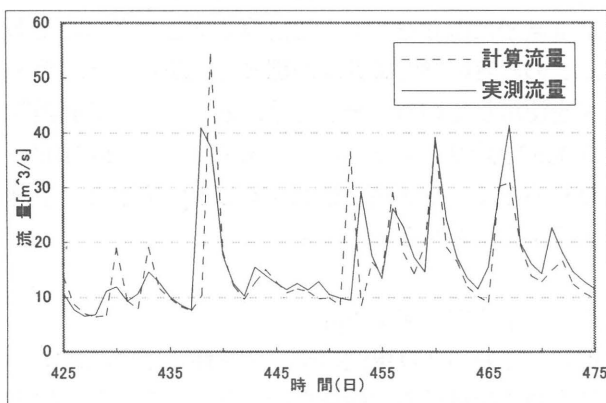


図13-2 日単位での予測結果(一部拡大)

5. 自己生成入力データによる肝属川流量の長期間予測の結果と考察

流量予測モデルが出来上がったところでこのモデル

を使い自己生成入力データによる長期間流量予測を行った。流量予測の場合は入力側では流量と雨量を入力するのに対し、出力は流量だけである。このことより、流量予測モデルで自己生成入力データによる複数時刻分の予測を行う場合、雨量の入力はどのようにするのかということになるが、この場合は雨が降らないことを前提として予測を行う場合には0を入力し、気象台により雨量が予測できるとすればその値を入力として用いることが考えられる。

現在の雨量予測のレベルは、時間単位の雨量予測では精度の高い予測ができるのは1時間後までとされているので、本稿では今後雨量予測の技術の発展が考えられること、及びモデルの有効性を検証するため、予測時には実際の雨量データを入力して検証することにした。各層の素子数などの予測条件は、これまでに求めていたパラメータを使用し、予測時の雨量データは前述したように実測雨量データを入力して予測を行っている。図15-1はデータの500時刻目から50時間予測を行った場合の予測結果、図15-2は525時刻目から50時間予測を行った場合の予測結果である。

図15-1は、雨量と流量ともに比較的少なく安定した状態から激しく増加してまた減少していく部分の予測だが、やはりこのように状態が急激に変化するような場合を長時間にわたり予測するのは難しいことが分かる。一方、図15-2は図15-1の後半部分にあたる525時刻目から50時間の予測であるが12時間分までは予測できているのに対し、537時刻目で実測流量よりも流量を大きく予測してしまい、その後の予測流量もその影響を受け、実測流量よりも大きな値になってしまっていることがはっきりとみてとれる。

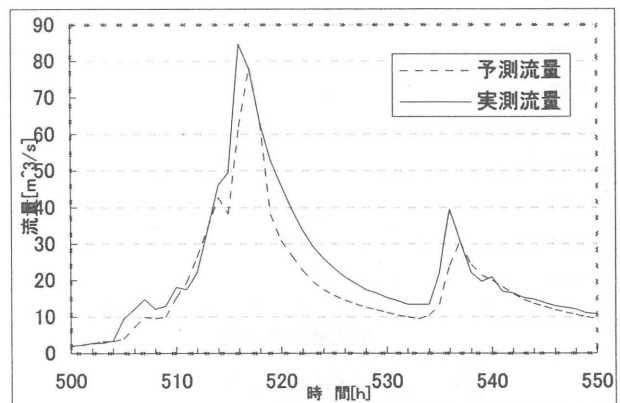


図15-1 予測結果1

この予測手法では自分の予測したデータをさらに次の予測に使用するという性質上、誤差の影響が蓄積していくため先の時間の予測になるほどその予測値の信頼度は低くなっていく。しかし従来の方式のモデルで

は出力層を増やすことでしか予測時間を増やすことはできず、5時間予測可能な出力層が5個のネットワークでさえほとんど使い物にならない結果しか得られなかったが、この自己生成入力データによる予測手法では、出力層素子数を増やすものに比べ、出力層は1つなので4-2の(2)で述べたような学習における問題も無い。また予測値を入力に戻す動作を繰り返すたびにいくらかでも先の時間の予測を行うことができるという利点がある。

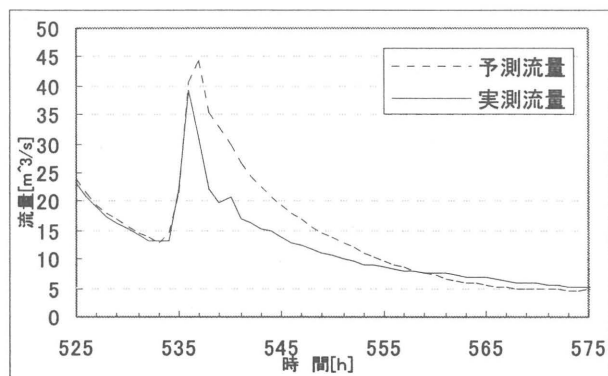


図 15-2 予測結果 2

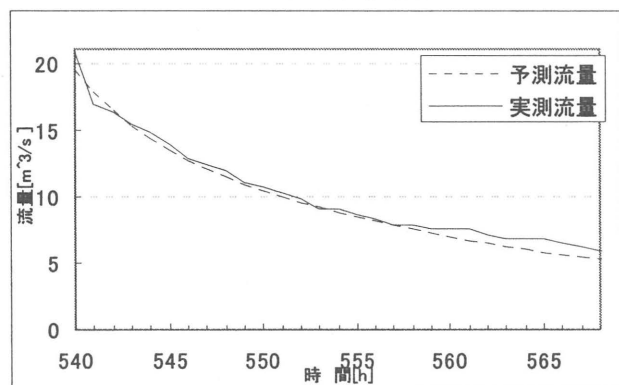


図 15-3 予測結果 3

さらに図 15-3 は雨がまったく降っていなかった期間、つまり自己生成による流量データのみで予測を行った結果だが、この結果と図 15-1 や図 15-2 を比較すると、このモデルでは雨量流量ともに増加する期間よりも雨が降らずに流量が減少していく期間の予測の方がよい予測結果を得られていることが分かる。

5. まとめと今後の課題

本稿では、ニューラルネットワークによって得られた時系列データの予測値を入力側に戻して使用することによりさらなる予測値を得るという自己生成入力データによる時系列データの長期間予測を行った。関数を使用して検証を行った結果、長期間にわたってよい予測値が得られており、肝属川の流量予測においても数十時間に及ぶ良好な予測値を得ることが出来た。

流量予測においては、データに測定誤差などが含ま

れていることもあり、値の確かな単純な周期関数の予測を比較して、精度の高い予測値を得ることは出来なかった。また予測値を自己生成入力データ方式で入力として使用してもその予測値の持っている誤差の影響を強く引き継ぐことなく次の予測値が得られている。その結果、長期間の流量の予測を行うことが出来た。

しかし流量が減っていく場合の予測は比較的うまくいっているが一番予測として重要な流量、雨量が急激に増加していく場合の予測精度があまりよくなかった。今後、入力にさらに周辺の雨量観測所のデータを取り入れるなどして改善を行う必要がある。また、流量予測モデルだけでなくほかの時系列データの長時間予測に自己生成入力データによる予測手法を取り入れ、その有効性を検証し、出来るだけ長い期間で予測精度を保った予測を行えるよう、更なる工夫をしていくことが必要である。

6. 謝辞

最後に資料収集にご協力いただいた本校土木工学科 足田教授と建設省大隈工事事務所に感謝の意を表します。

<参考文献>

- 1) 阿部清明, 菊池英明, 古川浩平, 塩月善晴: “ニューラルネットワークによる流出解析手法 (日流量) に関する研究” 土木学会論文集 No.656/II-52 (2000) 1-13
- 2) 飯坂達也, 松井哲郎, 植木芳照: “ニューロ・ファジーによるダム流入量予測システムの開発” 電学論 B 119 [10] (1999) 1020-1025
- 3) 大崎洋一郎, 永野俊: “経済指標の長期予測を行うニューラルネット予測システム” 電子情報通信学会論文誌 D-II Vol.J77-D-II No.2 (1994) 405-412
- 4) 須貝康雄, 堀部浩, 川瀬太郎: “基準需要を利用したニューラルネットによる翌日最大電力需要予測” 電学論 B 117 [6] (1997) 872-879
- 5) 水野勝教 ほか: “ニューラルネットワークを活用した地上雨量分布推定に用いる入力情報の検討” 電学論 B 120 [5] (2000) 665-671
- 6) 熊沢逸夫: 「学習とニューラルネットワーク」 森北出版 (1998) 1-77
- 7) 松葉育男: 「ニューラルシステムによる情報処理」 昭晃堂 (1993) 121-132
- 8) 建設省大隈工事事務所: 資料 (1993-1995)